Expanding the Hubway Network

Paula A. Hernandez

April 2018

1 Introduction

With the objective of being an accessible city, Boston is part of a public transportation system that connects it to around 175 cities and towns. Currently consisting of 145 subway stations, 177 bus routes, 8 ferry terminals, and 193 Hubway Stations (bike sharing) around the Greater Boston area, the public transportation network is constantly looking to expand its services.

In one of the City's most current efforts, residents are being asked to vote on proposed Hubway station locations. Motivated by this, I have decided to explore a computational way in which to expand the area covered by the Hubway Network, considering the distribution of houses around the City of Boston.

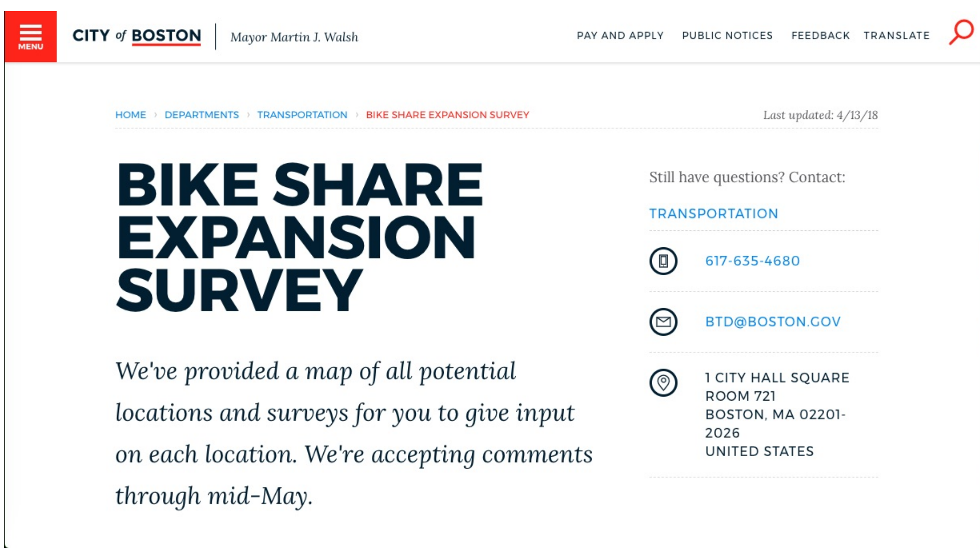

Figure 1: The Survey

2 Data Sets

Two data sets were retrieved from online and served as a basis for two more data sets that were later programmatically derived.

2.1 Online Data

A **Live Street Addresses Management (SAM) Addresses** data set was obtained from "Analyze Boston" portal, the data repository for the City of Boston. It contained **388,358 observations and 27 variables**. Each observation corresponds to a home address within the City of Boston (as opposed to within the Greater Boston Area). Out of all of these variables, only the Latitude and Longitude were of interest.

The second data set, **Hubway Stations as of July 2017**, was obtained from the Hubway System Data portal. This data set contained **193 observations and 6 variables**. Each of these observations corresponds to an existing Hubway Station. Again, the only two variables of interest were Latitude and Longitude.

2.2 Derived Data

One of the additional data sets, **Hubway Stations Home Density (Hubway DS)**, was derived from both of the above mentioned data sets. Each Hubway Station from the Hubway database was assigned a Home Density metric. Home Density was calculated to be the number of home addresses within a 1km radius centered at the Hubway Station for which that metric was being calculated. The reasoning for this calculation is that the homes within walking distance of each Hubway Station serve as a proxy for its demand. Walking distance for an average American is assumed to be 1km.

The last of the data sets, **Possible Network Extension Area (Possible Network)**, is also derived from both the SAM Addresses and Hubway Stations as of July data sets. This data set contains the addresses (as Latitude and Longitude tuples) that are more than 1km away from any Hubway Station and are within a 4km distance of the closest Hubway Station. This generates a 3km wide "ring", starting 1km away from the Hubway Stations that are in the edge of the Network. The reason for this data set is to identify the addresses that are not currently served by the Hubway Network, as they are not within walking distance from any Hubway Station (1km). At the same time, the addresses included in the data set are those which could be covered with the introduction of a new Hubway Station (4km away from the nearest Hubway Station). Any Hubway Station placed within this ring would result in an expansion of the area of coverage of the Network. This distance constraint for new Hubway Stations is put in place because of the Hubway pricing scheme: the first 30 minutes (3km) of riding time are free as result of the membership and any time over is charged at a premium. Thus, it follows that all Hubway Stations be placed within 30 minutes riding distance from one another.

3 Algorithms and Analysis Techniques

I believe that the problem of where to best place new Hubway Stations, so that they expand the area of coverage of the Network, is best modeled as an optimization problem. The addresses in the Possible Network data set are clustered using the k-means algorithm. After the clustering, each mean is assigned a Home Density metric (analogous to the one assigned to existing Hubway Stations). The means whose Home Density exceeds the average Home Density of the existing Hubway Stations are selected as possible locations for new Hubway Stations.

K-means ran with 5 means (found to be optimal considering the curve of Sum of Squared Errors). I use Home Density in order to determine if a mean found is a good location for a Hubway Station. This is backed by finding that Hubway Stations are placed in locations that have a high Home Density (high demand): there is a high, positive correlation between Home Density and number of Hubway Stations in a neighborhood.

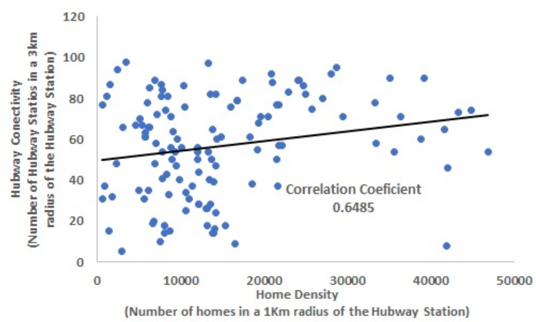

Figure 2: Home Density and Hubway Conectivity per Station

4 Visualizations and Web Application

To accompany the analysis above, I have created a web application that allows users to visualize the current Hubway Stations and the Proposed Stations on a map. On the map, the blue circles are the existing Hubway Stations and the red pins are the Proposed Stations. The web application also allows users to input an address and find out whether the address is within walking distance of a Hubway Station or a Proposed Hubway Station, within Network Expansion distance, or outside of both.

5 Limitations and Improvements

I identified at least two ways in which my analysis could be improved. These improvement should be incorporated in future versions of this project.

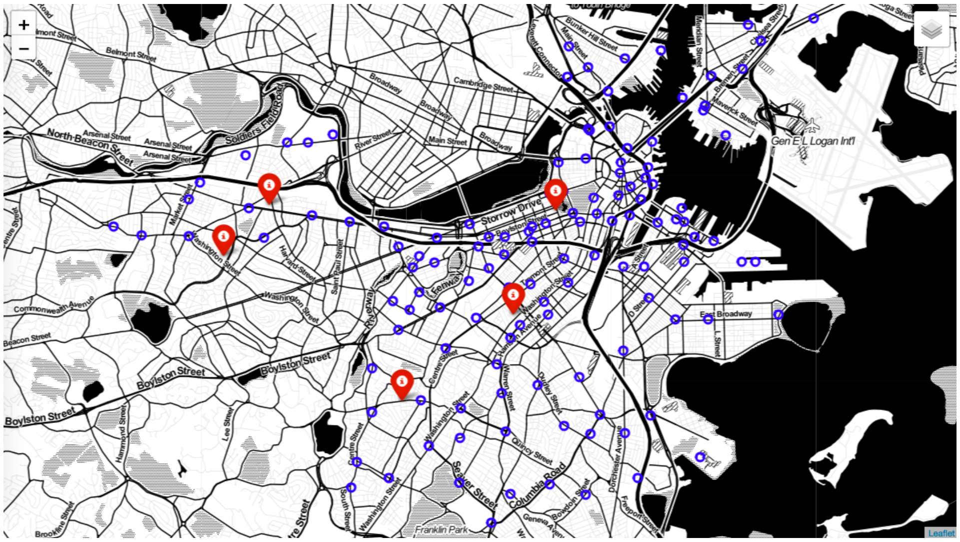

Figure 3: Map showing existing Hubway Stations (blue) and Proposed Stations (red)

- **Better Clustering:** A more accurate way of identifying possible locations for new Hubway Stations would use Home Density as the metric used to cluster the data in Possible Network. This would be preferred to using spatial distance as Home Density is the metric of interest that we consider when determining if a mean is a viable location for a new Hubway Station.

- **Stronger Proxy for Demand:** Hubway trip data could be incorporated to identify characteristics of stations that see more daily demand. Perhaps this information will reinforce my usage of Home Density as a proxy for Hubway demand or identify a better metric to be used in the search for good locations for new Hubway Stations.

Neither of these aspects were introduced to the analysis at hand due to time and computational constraints.

6 Conclusion

As the City of Boston seeks to expand the area covered by its public transportation services, it should seek a data driven plan. In regards to the Hubway Network, looking at characteristics that drive "current demand" and using those to find locations where "potential demand" is satisfactory for a new Hubway Station should be the approach taken. I have been able to identify 5 means which satisfy all of the requirements to be a good new location for a Hubway Station.